

AD-R173 498

NONPARAMETRIC INFERENCE IN ADDITIVE RISK MODELS FOR  
COUNTING PROCESSES..(U) FLORIDA STATE UNIV TALLAHASSEE  
DEPT OF STATISTICS I W MCKEAGUE AUG 86

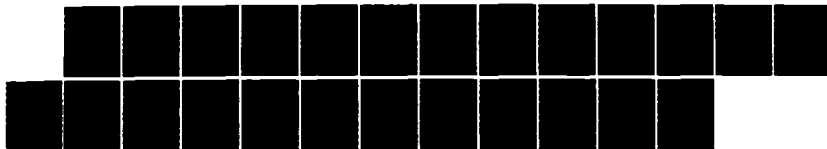
1/1

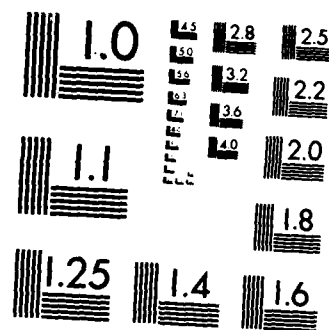
UNCLASSIFIED

FSU-STATISTICS-M741 ARO-23699. 8-MA

F/G 12/1

NL





MICROCOPY RESOLUTION TEST CHART  
NATIONAL BUREAU OF STANDARDS-1963-A

**OTIC FILE COPY**

②

NONPARAMETRIC INFERENCE IN ADDITIVE RISK

MODELS FOR COUNTING PROCESSES

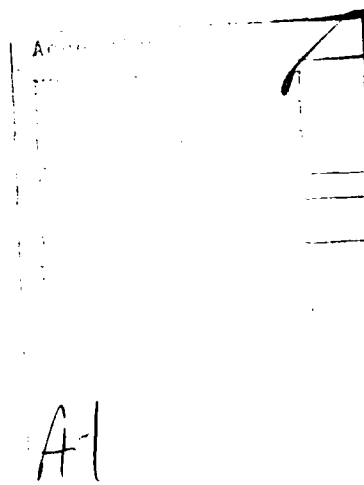
by

Ian W. McKeague

ABSTRACT

Nonparametric estimators for the hazard functions in an additive risk model for counting processes are studied. <sup>The hazard function</sup> We establish a functional central limit theorem for the integrated estimators and show how this can be used to find the asymptotic null distribution of a maximal deviation statistic for Kolmogorov-Smirnov type testing. <sup>for the</sup> In addition, we provide confidence bands for approximations to the integrated hazard functions and show that certain smoothed versions of the hazard function estimators are uniformly consistent.

*Keywords: nonparametric methods; regression models*



## 1. Introduction.

The proportional hazards regression model of Cox (1972) for the analysis of censored survival data has had considerable influence on the theory and practice of biostatistics. This model can be described as follows. Let  $\lambda(t) = \lambda\{t; Y(t)\}$  denote the "failure" rate at time  $t$  for a subject with covariate history  $Y(t) = (Y_1(t), \dots, Y_p(t))'$ , i.e.  $\lambda(t)dt$  is the probability that the subject dies in the small time interval from  $t$  to  $t + dt$ . Cox's model assumes that  $\lambda(t)$  has the multiplicative form

$$\lambda(t) = \alpha_0(t) \exp\left\{\sum_{j=1}^p \beta_j Y_j(t)\right\}, \quad (1.1)$$

where  $\beta_1, \dots, \beta_p$  are constants to be estimated and  $\alpha_0$  is an unknown baseline hazard rate.

An alternative model, introduced by Aalen (1980), is the additive risk model given by

$$\lambda(t) = \sum_{j=1}^p \alpha_j(t) Y_j(t), \quad (1.2)$$

where  $\alpha_1, \dots, \alpha_p$  are unknown "hazard" functions. Although statistical methods for Cox's model are well developed (e.g. Andersen and Gill, 1982), this has not been so for Aalen's model, except in the case of one covariate (Aalen, 1978). The object of the present paper is to provide techniques which can deal with any number of covariates in the Aalen model.

Aalen's model provides a useful alternative to the Cox model when the sample size is large and more detailed information concerning the influence of each covariate is needed. It could be used in epidemiologic studies involving large cohorts of individuals, such as those studies cited by Breslow (1986, pp. 110, 111), and in situations where the proportional hazards assumption of the Cox model is

poorly satisfied. Methods for assessing the goodness of fit of the Cox model are discussed by Arjas (1986).

Aalen (1975,1978,1980) formulated his model in the framework of counting processes as follows. Let  $X(t)$ ,  $t \in [0,1]$  be a counting process which counts observed events in the life of the subject observed over the time interval  $[0,1]$ . So the sample paths of  $X$  are step functions, zero at time zero, right-continuous with unit jumps. We assume that  $X$  has random intensity process  $\lambda(t)$  given by (1.2), where  $Y_1, \dots, Y_p$  are predictable covariate processes. By "intensity" we mean that the process  $M(t) = X(t) - \int_0^t \lambda(s)ds$  is a square integrable martingale. Writing  $X$  in differential form,  $dX(t) = \lambda(t)dt + dM(t)$ , we may regard  $dM(t)$  as additive "noise." The basic model equation is

$$X(t) = \int_0^t \lambda(s)ds + M(t), \quad (1.3)$$

where  $\lambda$  is given by (1.2).

The statistical problem is one of estimating the hazard functions  $\alpha_1, \dots, \alpha_p$  on the basis of  $n$  iid copies of  $X, Y_1, \dots, Y_p$  observed over  $[0,1]$ .

Our initial approach is to treat the hazard functions  $\alpha_1, \dots, \alpha_p$  as though they are piecewise constant functions. This reduces the problem to the estimation of finitely many parameters. It would be possible to use the method of maximum likelihood at this stage, as has been done by Buckley (1984). However, we are interested in obtaining asymptotic results which do not implicitly assume that the hazard functions are of a piecewise constant form. Under these circumstances the mathematics of maximum likelihood becomes intractable, except in the case of one covariate (Karr, 1983). A way out of this difficulty is to use a quasi-least-squares estimator for the parameters in the piecewise constant hazard functions. Quasi-least-squares estimators for this model were introduced by McKeague (1986a), and in a related model by Christopeit (1986) and Le Breton and Musiela (1986).

The final step is to allow the mesh size of the piecewise constant functions to tend to zero at a rate tied to the sample size  $n$ , obtaining asymptotic results as  $n \rightarrow \infty$ . This amounts to a use of the method of sieves (Grenander, 1981) involving the so-called histogram sieve. Let  $\hat{\alpha}_j^{(n)}(t)$  denote the estimator of  $\alpha_j$  obtained in this way. We shall obtain a functional central limit theorem for  $\hat{A}_j^{(n)}(t) = \int_0^t \hat{\alpha}_j^{(n)}(s) ds$ . This result is used to find the asymptotic distribution of a maximal deviation statistic for testing the hypothesis  $H_0: \alpha_j = \alpha_0$ , where  $\alpha_0$  is a known function. In addition, we shall provide confidence bands for the best approximation to the integrated hazard function  $A_j(t) = \int_0^t \alpha_j(s) ds$  within the histogram sieve and provide a uniformly consistent estimator for  $\alpha_j$ .

Our results are not restricted to counting processes. Indeed, processes satisfying equation (1.3) and a technical condition known as left-quasi-continuity are included. We have previously treated the functional central limit theorem for  $\hat{A}_j^{(n)}$  in the case that  $X$  is a continuous process (McKeague, 1986b) using orthogonal series sieves. An  $L^2$ -consistency result for  $\hat{\alpha}_j^{(n)}$  was established in McKeague (1986a).

## 2. Histogram sieve estimators.

$(\Omega, \mathcal{F}, P)$  will denote a complete probability space and  $(\mathcal{F}_t, t \in [0, 1])$  a nondecreasing right-continuous family of sub- $\sigma$ -fields of  $\mathcal{F}$  where  $\mathcal{F}_0$  contains all  $P$ -null sets in  $\mathcal{F}$ . All processes are indexed by  $t \in [0, 1]$ . The process  $M = (M(t), \mathcal{F}_t)$  is assumed to be a square integrable martingale such that almost all paths of  $M$  are right-continuous on  $[0, 1)$  with left limits on  $(0, 1]$ . Write  $\Delta M_t = M_t - M_{t-}$ , the jump in  $M$  at time  $t$ . A stochastic process  $X = (X(t), t \in [0, 1])$  is said to be left-quasi-continuous (see Dellacherie, 1972, p. 85) if for all predictable stopping times  $\tau$  taking values in  $[0, 1]$ ,  $X_\tau = X_{\tau-}$  almost surely. Here a stopping time  $\tau$  is said to be predictable if there is an increasing sequence  $(\tau_n)$  of stopping times such that  $\tau_n < \tau$  a.s. for all  $n \geq 1$  and  $\tau_n \uparrow \tau$  a.s. If  $X$  is a counting process with a continuous

compensator then it is left-quasi-continuous, see Liptser and Shirayev (1978, p. 243).

Let  $(X_i, M_i, Y_{ij}, j=1, \dots, p)$ ,  $i = 1, \dots, n$  denote  $n$  independent copies of the generic processes  $X$ ,  $M$  and  $Y_j$ ,  $j = 1, \dots, p$  which satisfy the model (1.3), where  $\lambda$  given by (1.2) represents the intensity in the case that  $X$  is a counting process. Let  $(d_n)$  be an increasing sequence of positive integers and define the histogram sieve estimator  $\hat{\alpha}_j^{(n)}$  by

$$\hat{\alpha}_j^{(n)}(t) = \beta_{rj} \quad \text{if } t \in \left[ \frac{r-1}{d_n}, \frac{r}{d_n} \right),$$

where the vector  $(\beta_{r1}, \dots, \beta_{rp})' \equiv \beta_r$  is given by

$$\beta_r = D_r^{-1} C_r, \quad (2.1)$$

$D_r$  is a  $p \times p$  matrix with entries

$$D_{rjk} = \sum_{i=1}^n \int_{(r-1)/d_n}^{r/d_n} Y_{ij}(s) Y_{ik}(s) ds, \quad (2.2)$$

and  $C_r$  is a column vector with components

$$C_{rj} = \sum_{i=1}^n \int_{(r-1)/d_n}^{r/d_n} Y_{ij}(s) dX_i(s). \quad (2.3)$$

Note that the computation of the histogram sieve estimator involves the inversion of  $d_n \times p$  matrices  $D_r$ ,  $r = 1, \dots, d_n$  (if  $D_r$  is singular then a generalized inverse will do) in contrast to the inversion of a single much larger  $pd_n \times pd_n$  matrix in the computation of orthogonal series sieve estimators studied by McKeague (1986a, 1986b).

The estimator  $\hat{\alpha}_j^{(n)}$  belongs to the set of functions  $S_{d_n} = \{\alpha: \alpha \text{ is constant on } [\frac{r-1}{d_n}, \frac{r}{d_n}), r=1, \dots, d_n\}$  and the sequence  $S_{d_n}$ ,  $n \geq 1$  is called the histogram sieve. The best approximation to  $\alpha_j$  (in the  $L^2[0,1]$  sense) by a member of  $S_{d_n}$  is given by

$$\alpha_j^{(n)}(t) = d_n \int_{(r-1)/d_n}^{r/d_n} \alpha_j(s) ds \quad \text{if } t \in \left[ \frac{r-1}{d_n}, \frac{r}{d_n} \right). \quad (2.4)$$



Write  $A_j^{\{n\}}(t) = \int_0^t \alpha_j^{\{n\}}(s) ds$ . A function  $f$  is said to be Lipschitz of order  $\gamma$ , where  $0 < \gamma \leq 1$ , if there is a constant  $C$  such that for all  $s, t$  in the domain of  $f$ ,  $|f(t) - f(s)| \leq C|t - s|^\gamma$ . If  $f$  is Lipschitz of order 1 we simply say it is Lipschitz.

We first state our functional central limit theorem for  $\hat{A}_j^{(n)}(t) = \int_0^t \hat{\alpha}_j^{(n)}(s) ds$  in the counting process case.

**Theorem 2.1.** Suppose that  $X$  is a counting process, the histogram sieve is used, conditions (C1) - (C4) (stated below) hold and  $d_n \uparrow \infty$ ,  $d_n = o(n^{1/2})$ . Then

$$\sqrt{n} (\hat{A}_j^{(n)} - A_j^{\{n\}}) \xrightarrow{D} m_j \text{ in } C[0,1],$$

where  $m_j$  is a continuous Gaussian martingale with mean zero and covariance function

$$\text{Cov}(m_j(s), m_j(t)) = \int_0^{s \wedge t} L_j^2(u) \left\{ \sum_{k=1}^p \alpha_k(u) E[Y_j^2(u) Y_k(u)] \right\} du,$$

where  $L_j(u)$  is the  $j^{\text{th}}$  element of  $\text{diag}[K^{-1}(u)]$ , in which  $K^{-1}(u)$  is the inverse of the  $p \times p$  matrix  $K(u)$  having components  $K_{r\ell}(u) = E[Y_r(u) Y_\ell(u)]$ .

#### Conditions.

(C1)  $\alpha_j$  is Lipschitz of order  $\gamma > \frac{1}{2}$  for  $j = 1, \dots, p$ .

(C2)  $\inf_{t \in [0,1]} E Y_j^2(t) > 0$  for  $j = 1, \dots, p$ .

(C3) 
$$\frac{\sup_{t \in [0,1]} |E Y_j(t) Y_k(t)|}{\left[ \inf_{t \in [0,1]} E Y_j^2(t) \right]^{1/2} \left[ \inf_{t \in [0,1]} E Y_k^2(t) \right]^{1/2}} < \frac{1}{p-1}$$

for all  $1 \leq j < k \leq p$ , applicable for  $p \geq 2$ .

(C4)  $\sup_{t \in [0,1]} E |Y_j(t)|^5 < \infty$  for  $j = 1, \dots, p$ .

Remarks.

(i) In the special case of a single covariate ( $p=1$ ) we have that the asymptotic variance of the estimator  $\hat{A}_1^{(n)}(t)$  is

$$Em_1^2(t) = \int_0^t \frac{\alpha_1(s) EY_1^3(s)}{[EY_1^2(s)]^2} ds. \quad (2.5)$$

However, a natural estimator for  $A_1$  in the case  $p=1$  is the well known Nelson-Aalen estimator given by

$$\tilde{A}_1^{(n)}(t) = \int_0^t \frac{dX^{(n)}(s)}{Y^{(n)}(s)}, \quad (2.6)$$

where  $X^{(n)}(t) = \sum_{i=1}^n X_i(t)$  and  $Y^{(n)}(t) = \sum_{i=1}^n Y_{i1}(t)$ . Here  $1/0$  is defined to be zero.

Aalen (1978) showed that  $\sqrt{n}(\tilde{A}_1^{(n)} - A_1)$  converges weakly in the Skorohod space  $D[0,1]$  to a continuous Gaussian martingale  $m_0$  such that

$$Em_0^2(t) = \int_0^t \frac{\alpha_1(s)}{EY_1(s)} ds. \quad (2.7)$$

It is interesting to compare (2.5) and (2.7). Note that  $Em_0^2(t) \leq Em_1^2(t)$  by Hölder's inequality. Aalen (1980) has extended  $\tilde{A}_1^{(n)}$  to give estimators  $\tilde{A}_1^{(n)}, \dots, \tilde{A}_p^{(n)}$  of  $A_1, \dots, A_p$  when  $p > 1$ , however it has not been possible to obtain consistency or asymptotic normality results for these estimators.

(ii) McKeague (1986b) showed that (C3) implies  $K(t)$  is non-singular for almost every  $t$ , so the covariance function of  $m_j$  in the statement of Theorem 2.1 is well defined. It seems reasonable to conjecture that Theorem 2.1 remains true when the quite restrictive condition (C3) is replaced by the condition that the minimum eigenvalue of  $K(t)$  is bounded away from zero. We could have replaced (C3) by a weak condition of this kind, but to avoid technical distractions we have refrained from doing so. From the point of view of applications it would be safe to disregard condition (C3).

Theorem 2.1 is a consequence of the following more general result.  $\langle M \rangle$  denotes the predictable quadratic variation process of  $M$ , i.e. the unique increasing predictable process such that  $M_t^2 - \langle M \rangle_t$  is a martingale.

Theorem 2.2. Suppose that  $X$  is a left-quasi-continuous process, the histogram sieve is used, conditions (C1) - (C3), (D1) - (D5) (stated below) hold and  $d_n \uparrow \infty$ ,  $d_n = o(n^{1/2})$ . Then

$$\sqrt{n}(\hat{A}_j^{(n)} - A_j^{\{n\}}) \xrightarrow{\mathcal{D}} m_j \quad \text{in } C[0,1],$$

where  $m_j$  is a continuous Gaussian martingale with mean zero and covariance function

$$\text{Cov}(m_j(s), m_j(t)) = E \int_0^{s \wedge t} L_j^2(u) Y_j^2(u) d\langle M \rangle_u,$$

where  $L_j$  is defined in the statement of Theorem 2.1.

#### Conditions

$$(D1) \quad \sup_{t \in [0,1]} E Y_j^4(t) < \infty \quad \text{for } j = 1, \dots, p.$$

(D2) The predictable variation process  $\langle M \rangle$  has absolutely continuous sample paths (a.s.) and there exists  $\gamma > 1$  such that

$$\sup_{t \in [0,1]} E \left[ |Y_j(t)|^{2\gamma} \left( \frac{d\langle M \rangle_t}{dt} \right)^\gamma \right] < \infty, \quad \text{for } j = 1, \dots, p.$$

(D3) The process  $\sum_{0 < s \leq t} (\Delta M_s)^2$  and its compensator have finite second moments at  $t = 1$ .

$$(D4) \quad E \left( \sum_{0 < s \leq 1} Y_j^2(s) (\Delta M_s)^2 \right) < \infty \quad \text{for } j = 1, \dots, p.$$

(D5) The function

$$\delta_j(t) = E \left[ \int_0^t Y_j^4(s) d\pi_s \right], \quad t \in [0,1]$$

where  $\pi = (\pi_s)$  is the compensator of the process  $\sum_{0 < s \leq t} (\Delta M_s)^4$  is Lipschitz for  $j = 1, \dots, p$ .

### 3. The maximal deviation statistic.

Define  $p \times p$  matrices  $\hat{K}^{(n)}$  and  $\hat{M}^{(n)}$  by

$$\hat{K}_{jk}^{(n)}(t) = n^{-1} \sum_{i=1}^n Y_{ij}(t) Y_{ik}(t),$$

$$\hat{M}_{jk}^{(n)}(t) = n^{-1} \sum_{i=1}^n Y_{ij}^2(t) Y_{ik}(t),$$

and let  $\hat{L}_j^{(n)}(t)$  be the  $j^{\text{th}}$  element of the diagonal of a generalized inverse of  $\hat{K}^{(n)}$ .

In order to test the hypothesis  $H_0: \alpha_j = \alpha_0$ , where  $\alpha_0$  is a known function, we propose the use of the statistic

$$T_j^{(n)} = \sqrt{n} \hat{G}_j^{(n)}(1)^{1/2} \sup_{t \in [0,1]} \left| \frac{\hat{A}_j^{(n)}(t) - A_0^{(n)}(t)}{\hat{G}_j^{(n)}(1) + \hat{G}_j^{(n)}(t)} \right|,$$

where  $A_0^{(n)}(t) = \int_0^t \alpha_0^{(n)}(s) ds$ ,  $\alpha_0^{(n)}$  is the best  $L^2$ -approximation to  $\alpha_0$  from within  $S_{d_n}$  and

$$\hat{G}_j^{(n)}(t) = \int_0^t [\hat{L}_j^{(n)}(s)]^2 \{ \alpha_0(s) \hat{M}_{jj}^{(n)}(s) + \sum_{k \neq j} \alpha_k^{(n)}(s) \hat{M}_{jk}^{(n)}(s) \} ds.$$

The following result gives the asymptotic null distribution of  $T_j^{(n)}$ .

**Theorem 3.1.** Suppose that the conditions of Theorem 2.1 are satisfied, the sample paths of  $Y_k$  are left continuous with right hand limits,  $\inf_{t \in [0,1]} |\det K(t)| > 0$  and for  $k = 1, \dots, p$ ,

$$E \left[ \sup_{t \in [0,1]} |Y_k(t)|^3 \right] < \infty. \quad (3.1)$$

Then if  $H_0$  holds

$$\lim_{n \rightarrow \infty} P\{T_j^{(n)} \geq c_\alpha\} = \alpha,$$

where  $0 < \alpha < 1$ ,  $c_\alpha$  is the upper  $\alpha$  quantile of the distribution of  $\sup_{t \in [0, 1/2]} |B^0(t)|$  and  $B^0$  is the Brownian bridge process.

Using arguments in the proof of Theorem 3.1 it can be checked that  $T_j^{(n)}$  provides a consistent test against all alternatives in the sense that  $T_j^{(n)} \rightarrow \infty$  a.s. under any alternative. A table for the distribution of  $\sup_{t \in [0, \frac{1}{2}]} |B^0(t)|$  has been given by Hall and Wellner (1980). For instance  $c_{.05} = 1.273$ .

#### 4. Confidence bands.

Under the condition

$$\sqrt{n} \sup_{t \in [0, 1]} |A_j(t) - A_j^{\{n\}}(t)| \rightarrow 0 \quad (4.1)$$

Theorem 2.1 holds with  $A_j^{\{n\}}$  replaced by  $A_j$ . So under (4.1) we may obtain confidence bands for  $A_j$ . Unfortunately the class of functions  $\alpha_j$  for which (4.1) holds with  $d_n = o(n^{\frac{1}{2}})$  is too small to be of practical interest. We are obliged to make do with confidence bands for  $A_j^{\{n\}}$ .

First estimate  $G_j(t) = E m_j^2(t)$  by  $\hat{G}_j^{(n)}(t)$ , as in section 3 but with  $\alpha_0$  replaced by  $\hat{\alpha}_j^{(n)}$ . Then under the conditions of Theorem 3.1 we see that an asymptotic  $100(1-\alpha)\%$  confidence band for  $A_j^{\{n\}}$  has upper and lower limits given by

$$\hat{A}_j^{(n)}(t) \pm c_\alpha n^{-\frac{1}{2}} \hat{G}_j^{(n)}(1)^{\frac{1}{2}} \left(1 + \frac{\hat{G}_j^{(n)}(t)}{\hat{G}_j^{(n)}(1)}\right), \quad t \in [0, 1]$$

where  $c_\alpha$  is defined in the statement of Theorem 3.1.

#### 5. Kernel estimators for the hazard functions.

The functions of real interest are the hazard functions  $\alpha_1, \dots, \alpha_p$  rather than the integrated hazard functions  $A_1, \dots, A_p$ . It would be difficult to obtain an adequate picture of  $\alpha_j$  by visually assessing the gradient of  $\hat{A}_j^{(n)}$ . The estimator  $\hat{\alpha}_j^{(n)}$  itself has only been shown to be consistent in the  $L^2$ -sense:  $\int_0^1 [\hat{\alpha}_j^{(n)}(t) - \alpha_j(t)]^2 dt \xrightarrow{P} 0$ , see McKeague (1986a), and experience shows that it has very rough pathwise behaviour. However, in the following theorem we are

able to show that a smoothed version of  $\hat{\alpha}_j^{(n)}$  provides a uniformly consistent estimator of  $\alpha_j$ . This result was obtained under the assumption that  $X$  is a continuous process by McKeague (1986b, Theorem 3.2). Let  $K$  be a (kernel) function having integral 1 and support  $[-1,1]$ . Define

$$\tilde{\alpha}_j^{(n)}(t) = \frac{1}{b_n} \int_0^1 K\left(\frac{t-s}{b_n}\right) \hat{\alpha}_j^{(n)}(s) ds,$$

where  $b_n > 0$  is a bandwidth parameter.

Theorem 5.1. Suppose that  $X$  is a counting process, the histogram sieve is used, conditions (C1) - (C4) hold,  $\alpha_j$  and  $K$  are Lipschitz,  $b_n = n^{-\beta}$  where  $\frac{1}{4} \leq \beta < \frac{1}{2}$  and  $d_n = [n^\delta]$  where  $\frac{1}{2}(1-\beta) \leq \delta < \frac{1}{2}$ . Then

$$\sup_{t \in [0,1]} |\tilde{\alpha}_j^{(n)}(t) - \alpha_j(t)| = o_p(n^{-\frac{1}{2}(1-2\beta)}).$$

In future work we shall study the performance of the above estimators and test statistics using simulated data and discuss their application to real data.

## 6. Proofs.

The following lemma is needed for the proof of Theorem 2.2.

Lemma 6.1 Suppose that  $M$  is a left-quasi-continuous square integrable martingale satisfying condition (D3). Let  $2 \leq q \leq 4$ . Then there exist constants  $C_1, C_2$  such that for each predictable process  $(H_t, t \in [0,1])$  satisfying

$$E \int_0^1 H_t^2 d\langle M \rangle_t < \infty, \quad (6.1)$$

$$E \left( \sum_{0 < t \leq 1} H_t^2 (\Delta M_t)^2 \right) < \infty, \quad (6.2)$$

the following inequality holds:

$$E \left| \int_0^1 H_t dM_t \right|^q \leq C_1 E \left( \int_0^1 H_t^2 d\langle M \rangle_t \right)^{q/2} + C_2 (E \int_0^1 H_t^4 d\pi_t)^{q/4},$$

where  $\pi = (\pi_t)$  is the compensator of  $\sum_{0 < s \leq t} (\Delta M_s)^4$ .

Proof. Decompose  $M$  into the sum of a continuous martingale  $M^c$  and a purely discontinuous martingale  $M^d$  such that  $\langle M^c, M^d \rangle = 0$ , see Dellacherie and Meyer (1982, VIII 43).  $\langle M^d \rangle$  is the compensator of the square bracket process  $[M^d]_t = \sum_{0 < s \leq t} (\Delta M_s)^2$  which is left-quasi-continuous. Thus by Dellacherie (1972, p. 111),  $\langle M^d \rangle$  is continuous. The process  $Q \equiv [M] - \langle M \rangle = [M^d] - \langle M^d \rangle$  is a square integrable martingale (by (D3)),

$$[Q]_t = [[M^d] - \langle M^d \rangle] = [[M^d]] = \sum_{0 < s \leq t} (\Delta M_s)^4,$$

and  $\langle Q \rangle_t = \pi_t$  where  $\pi_t$  is defined in the statement of the Lemma. If  $E \int_0^1 H_t^4 d\langle Q \rangle_t = \infty$  nothing remains to be proved, so assume that  $E \int_0^1 H_t^4 d\langle Q \rangle_t < \infty$ . Then the stochastic integral  $\int_0^1 H_t^2 dQ_t$  is defined. Conditions (6.1) and (6.2) ensure that  $E \int_0^1 H_t^2 |dQ_t| < \infty$ , so by Kopp (1984, Theorem 4.3.18) the stochastic integral and Stieltjes integral interpretations of  $\int_0^1 H_t^2 dQ_t$  coincide. By the Burkholder-Davis-Gundy inequality (see Dellacherie and Meyer, 1982, p. 287) there exists a constant  $C$  such that

$$\begin{aligned} E \left| \int_0^1 H_t dM_t \right|^q &\leq C E \left( \int_0^1 H_t^2 d[M]_t \right)^{q/2} \\ &\leq C_1 E \left( \int_0^1 H_t^2 d\langle M \rangle_t \right)^{q/2} + C_2 E \left| \int_0^1 H_t^2 dQ_t \right|^{q/2}, \end{aligned}$$

where we have used  $|a+b| \leq 2^{\gamma-1}(|a|^\gamma + |b|^\gamma)$  for  $\gamma \geq 1$ , and  $C_1 = C_2 = 2^{\frac{q}{2}-1} C$ . Finally, by Lyapounov's inequality,

$$\begin{aligned} E \left| \int_0^1 H_t^2 dQ_t \right|^{q/2} &\leq [E \left( \int_0^1 H_t^2 dQ_t \right)^2]^{q/4} \\ &= [E \int_0^1 H_t^4 d\pi_t]^{q/4}, \end{aligned}$$

which completes the proof.  $\square$

Proof of Theorem 2.2. It is clear that for each  $n \geq 1$  the histogram sieve estimator  $\hat{\alpha}_j^{(n)}$  can be written in the form of the orthogonal series sieve estimator considered in McKeague (1986a, 1986b), where the orthonormal vectors  $\phi_{jr}$ ,  $r = 1, \dots, d_n$  used to define  $\hat{\alpha}_j^{(n)}$  are replaced by

$$\phi_r^{(n)}(t) = \begin{cases} d_n & \text{if } t \in [\frac{r-1}{d_n}, \frac{r}{d_n}) \\ 0 & \text{otherwise,} \end{cases}$$

$r = 1, \dots, d_n$ . Now a formal rewriting of the proof of Corollary 2.3 and Theorem 2.1 of McKeague (1986b), expanding  $\hat{\alpha}_j^{(n)}$  in terms of  $\phi_r^{(n)}$ ,  $r = 1, \dots, d_n$ , shows that the finite dimensional distributions of  $\sqrt{n}(\hat{A}_j^{(n)} - A_j^{(n)})$  converge weakly to those of  $m_j$ , the required conditions holding when (C1) - (C3), (D1) and (D2) are satisfied. Moreover, we have the representation

$$\sqrt{n}(\hat{A}_j^{(n)}(t) - A_j^{(n)}(t)) = U_n(t) + V_n(t), \quad (6.3)$$

where  $\sup_{t \in [0,1]} |U_n(t)| \xrightarrow{p} 0$ ,

$$\begin{aligned} V_n(t) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{ni}(t) \\ Z_{ni}(t) &= \sum_{k=1}^p \int_0^1 u_k^{(n)}(s, t) Y_{ik}(s) dM_i(s) \\ u_k^{(n)}(s, t) &= \sum_{r=1}^{d_n} \lambda_{kr}^{(n)}(t) \phi_r^{(n)}(s) \\ \text{vec}[\lambda^{(n)}(t)] &= R^{(n)-1} \text{vec}[h^{(n)}(t)] \\ h_{kr}^{(n)}(t) &= \begin{cases} \int_0^1 \mathbf{1}_{[0,t]}(s) \phi_r^{(n)}(s) ds & \text{for } k = j, r = 1, \dots, d_n \\ 0 & \text{for } k \neq j, r = 1, \dots, d_n, \end{cases} \end{aligned}$$

where  $R^{(n)}$  is the  $pd_n \times pd_n$  matrix partitioned into the  $p^2$  submatrices  $R_{km}^{(n)}$ ,  $k, m = 1, \dots, p$  with entries

$$R_{kmr\ell}^{(n)} = \int_0^1 \phi_r^{(n)}(t) \phi_\ell^{(n)}(t) E[Y_k(t) Y_m(t)] dt.$$



The proof of Lemma 4.3 of McKeague (1986b) shows that  $R^{(n)}$  is invertible for all  $n \geq 1$  and

$$\sup_{n \geq 1} \|R^{(n)-1}\| < \infty, \quad (6.4)$$

where  $\|\cdot\|$  denotes operator norm. From (6.3) it remains to show that the sequence of processes  $\{V_n, n \geq 1\}$  is tight in  $C[0,1]$ . If we can find constants  $q \geq 0$ ,  $\gamma > 1$  and  $C$  such that for all  $n \geq 1$ ,  $t_1, t_2 \in [0,1]$

$$E|V_n(t_2) - V_n(t_1)|^q \leq C|t_2 - t_1|^\gamma \quad (6.5)$$

then, by Billingsley (1968, Theorem 12.3),  $\{V_n, n \geq 1\}$  will be tight. In what follows  $C$  denotes a generic positive constant which is independent of  $n$ . By the Marcinkiewicz-Zygmund inequality (see Chow and Teicher, 1978, p. 356) for  $q \geq 1$

$$\begin{aligned} E|V_n(t_2) - V_n(t_1)|^q &\leq CE \left\{ \frac{1}{n} \sum_{i=1}^n [Z_{ni}(t_2) - Z_{ni}(t_1)]^2 \right\}^{q/2} \\ &\leq CE |Z_{n1}(t_2) - Z_{n1}(t_1)|^q \\ &\leq C \sum_{k=1}^p E \left| \int_0^1 \{u_k^{(n)}(s, t_2) - u_k^{(n)}(s, t_1)\} Y_k(s) dM_s \right|^q. \end{aligned} \quad (6.6)$$

Since  $\phi_r^{(n)}$  is a bounded function so is  $u_k^{(n)}(\cdot, t)$  (for each fixed  $t$ ) and it follows, using conditions (D2) and (D4), that for fixed  $k, n, t_1, t_2$  the predictable process  $H_s = v(s)Y_k(s)$ , where  $v(s) = u_k^{(n)}(s, t_2) - u_k^{(n)}(s, t_1)$ , satisfies conditions (6.1), (6.2) of Lemma 6.1. Thus, for  $2 \leq q \leq 4$

$$\begin{aligned} E \left| \int_0^1 \{u_k^{(n)}(s, t_2) - u_k^{(n)}(s, t_1)\} Y_k(s) dM_s \right|^q &\leq C_1 E \left( \int_0^1 v^2(s) Y_k^2(s) d\langle M \rangle_s \right)^{q/2} \\ &\quad + C_2 (E \int_0^1 v^4(s) Y_k^4(s) d\pi_s)^{q/4}. \end{aligned} \quad (6.7)$$

Now choose  $\gamma$  satisfying (D2) and  $1 < \gamma \leq 2$ . Then setting  $q = 2\gamma$  we have

$$\begin{aligned}
E \left( \int_0^1 v^2(s) Y_k^2(s) d\langle M \rangle_s \right)^{q/2} &\leq E \int_0^1 |v(s)|^q |Y_k(s)|^q \left( \frac{d\langle M \rangle}{ds} \right)^{q/2} ds \\
&\leq C \int_0^1 |v(s)|^q ds \quad (\text{by (D2)}) \\
&\leq C \left( \int_0^1 v^4(s) ds \right)^{q/4}
\end{aligned}$$

so that, using condition (D5) on the last term in (6.7),

$$\begin{aligned}
&E \left| \int_0^1 \{u_k^{(n)}(s, t_2) - u_k^{(n)}(s, t_1)\} Y_k(s) dM_s \right|^{2\gamma} \\
&\leq C \left( \int_0^1 \left| \sum_{r=1}^{d_n} [\lambda_{kr}^{(n)}(t_2) - \lambda_{kr}^{(n)}(t_1)] \phi_r^{(n)}(s) \right|^4 ds \right)^{\gamma/2}. \quad (6.8)
\end{aligned}$$

From the definition of  $\phi_r^{(n)}$  we note that for  $r_1, \dots, r_4 \leq d_n$ ,

$$\int_0^1 \phi_{r_1}^{(n)}(s) \phi_{r_2}^{(n)}(s) \phi_{r_3}^{(n)}(s) \phi_{r_4}^{(n)}(s) ds = \begin{cases} 1 & \text{if } r_1 = r_2 = r_3 = r_4 \\ 0 & \text{otherwise.} \end{cases}$$

It follows that

$$\begin{aligned}
\int_0^1 \left| \sum_{r=1}^{d_n} [\lambda_{kr}^{(n)}(t_2) - \lambda_{kr}^{(n)}(t_1)] \phi_r^{(n)}(s) \right|^4 ds &= \sum_{r=1}^{d_n} [\lambda_{kr}^{(n)}(t_2) - \lambda_{kr}^{(n)}(t_1)]^4 \\
&\leq \left\{ \sum_{r=1}^{d_n} [\lambda_{kr}^{(n)}(t_2) - \lambda_{kr}^{(n)}(t_1)]^2 \right\}^2 \\
&= \|\text{vec} [\lambda^{(n)}(t_2) - \lambda^{(n)}(t_1)]\|^4 \\
&\leq C \|\text{vec} [h^{(n)}(t_2) - h^{(n)}(t_1)]\|^4 \quad (\text{by (6.4)}) \\
&\leq C \left\{ \int_0^1 (1_{[0, t_2]}(s) - 1_{[0, t_1]}(s))^2 ds \right\}^2 \\
&\quad (\text{by Bessel's inequality}) \\
&\leq C |t_2 - t_1|^2.
\end{aligned}$$

Combining this with (6.8) shows that (6.5) is satisfied with  $q = 2\gamma$ ,  $\gamma > 1$ .  $\square$

Proof of Theorem 2.1. It suffices to check that the conditions of Theorem 2.2 are satisfied. By Liptser and Shirayev (1978, p. 243) the counting process  $X$  is left-quasi-continuous. (D1) is a consequence of (C4). Since  $\langle M \rangle_t = \int_0^t \lambda(s) ds = \int_0^t (\sum_{j=1}^p \alpha_j(s) Y_j(s)) ds$ , (D2) holds with  $\gamma = 2$  as a consequence of (C1) and (C4). In the counting process case  $\sum_{0 < s \leq t} (\Delta M_s)^2$  coincides with  $X(t)$  which has compensator  $\int_0^t \lambda(s) ds$  so that (D3) is implied by (C1) and (C4). For condition (D4),

$$\begin{aligned} E\left(\sum_{0 < s \leq 1} Y_j^2(s) (\Delta M_s)^2\right) &= E\left(\int_0^1 Y_j^2(s) dX(s)\right) \\ &= E\left(\int_0^1 Y_j^2(s) \lambda(s) ds\right) = \sum_{k=1}^p \int_0^1 \alpha_k(s) E[Y_j^2(s) Y_k(s)] ds < \infty, \end{aligned}$$

by (C1) and (C4).  $\square$

Proof of Theorem 3.1. Condition (3.1) implies that

$$E\left(\sup_{t \in [0,1]} |Y_j(t) Y_k(t)|\right) < \infty \quad (6.9)$$

and

$$E\left(\sup_{t \in [0,1]} |Y_j^2(t) Y_k(t)|\right) < \infty, \quad (6.10)$$

for  $j, k = 1, \dots, p$  using Hölder's inequality. Applying the strong law of large numbers in  $D[0,1]$  (Ranga Rao, 1963, Theorem 1) in the reversed time direction, given (6.9) and (6.10) it follows that for all  $j, k = 1, \dots, p$

$$\sup_{t \in [0,1]} |\hat{K}_{jk}^{(n)}(t) - K_{jk}(t)| \xrightarrow{\text{a.s.}} 0 \quad (6.11)$$

and

$$\sup_{t \in [0,1]} |\hat{M}_{jk}^{(n)}(t) - M_{jk}(t)| \xrightarrow{\text{a.s.}} 0, \quad (6.12)$$

where  $M_{jk}(t) = E[Y_j^2(t) Y_k(t)]$ . By (6.11) and the condition on  $\det \hat{K}(t)$ ,  $K^{(n)}(t)$  is nonsingular for all  $t \in [0,1]$  for  $n$  sufficiently large a.s. so by the

componentwise continuity of the matrix inverse operation we have

$$\sup_{t \in [0,1]} |\hat{L}_j^{(n)}(t) - L_j(t)| \xrightarrow{a.s.} 0. \quad (6.13)$$

Let  $G_j(t) = E m_j^2(t)$ , given in the statement of Theorem 2.1. Under the conditions of the Theorem,

$$\int_0^1 [\hat{\alpha}_j^{(n)}(t) - \alpha_j(t)]^2 dt \xrightarrow{P} 0, \quad (6.14)$$

from a histogram sieve version of Theorem 2.1 of McKeague (1986a). Using (6.12)-(6.14) and the Cauchy-Schwarz inequality it is then easy to show that

$$\sup_{t \in [0,1]} |\hat{G}_j^{(n)}(t) - G_j(t)| \xrightarrow{P} 0. \quad (6.15)$$

Let

$$Z_n(t) = \sqrt{n} G_j(1)^{\frac{1}{2}} \left[ \frac{\hat{A}_j^{(n)}(t) - A_0^{(n)}(t)}{G_j(1) + G_j(t)} \right]$$

and note that by Theorem 2.1  $Z_n$  converges weakly in  $C[0,1]$  to the process

$B^0 \left( \frac{G_j(t)}{G_j(1) + G_j(t)} \right)$ ,  $t \in [0,1]$ , where  $B^0$  is the Brownian bridge process. By the continuous mapping theorem (Billingsley, 1968, Theorem 5.1)  $\sup_{t \in [0,1]} |Z_n(t)|$  converges in distribution to  $\sup_{t \in [0, \frac{1}{2}]} |B^0(t)|$ . But

$$\begin{aligned} & \sup_{t \in [0,1]} \left| \sqrt{n} \hat{G}_j^{(n)}(1)^{\frac{1}{2}} \left[ \frac{\hat{A}_j^{(n)}(t) - A_0^{(n)}(t)}{\hat{G}_j^{(n)}(1) + \hat{G}_j^{(n)}(t)} \right] - Z_n(t) \right| \\ & \leq \sup_{t \in [0,1]} |Z_n(t)| \sup_{t \in [0,1]} \left| \frac{\hat{G}_j^{(n)}(1)^{\frac{1}{2}} (G_j(1) + G_j(t))}{G_j(1)^{\frac{1}{2}} (\hat{G}_j^{(n)}(1) + \hat{G}_j^{(n)}(t))} - 1 \right| \end{aligned}$$

which tends to 0 in probability since the first term is tight and (6.15) shows that the second term tends to zero in probability. The result now follows from Theorem 4.1 of Billingsley (1968).  $\square$

Proof of Theorem 5.1. Define the following smoothed versions of  $\alpha_j$  and  $\alpha_j^{\{n\}}$ :

$$\begin{aligned}\bar{\alpha}_j^{(n)}(t) &= \frac{1}{b_n} \int_0^1 K\left(\frac{t-s}{b_n}\right) \alpha_j(s) ds, \\ \alpha_j^{*(n)}(t) &= \frac{1}{b_n} \int_0^1 K\left(\frac{t-s}{b_n}\right) \alpha_j^{\{n\}}(s) ds.\end{aligned}$$

Since  $K$  is Lipschitz it is of bounded variation. Denote its total variation by  $V(K)$ . Then (c.f. the proof of Theorem 4.1.2. of Ramlau-Hansen(1983)),

$$\begin{aligned}\sup_{t \in [0,1]} |\bar{\alpha}_j^{(n)}(t) - \alpha_j^{*(n)}(t)| &= \sup_{t \in [0,1]} \left| \frac{1}{b_n} \int_0^1 K\left(\frac{t-s}{b_n}\right) d(\hat{A}_j^{(n)} - A_j^{\{n\}})(s) \right| \\ &\leq \frac{2}{b_n} V(K) \sup_{s \in [0,1]} |\hat{A}_j^{(n)}(s) - A_j^{\{n\}}(s)| \\ &= O_p\left(\frac{1}{b_n \sqrt{n}}\right),\end{aligned}\tag{6.16}$$

since  $\{\sqrt{n}(\hat{A}_j^{(n)} - A_j^{\{n\}}), n \geq 1\}$  is tight in  $C[0,1]$  by Theorem 2.1. Next

$$\begin{aligned}|\alpha_j^{*(n)}(t) - \bar{\alpha}_j^{(n)}(t)| &= \left| \frac{1}{b_n} \int_0^1 K\left(\frac{t-s}{b_n}\right) [\alpha_j^{\{n\}}(s) - \alpha_j(s)] ds \right| \\ &\leq \frac{1}{b_n} \left[ \int_0^1 K^2\left(\frac{t-s}{b_n}\right) ds \right]^{1/2} \left[ \int_0^1 [\alpha_j^{\{n\}}(s) - \alpha_j(s)]^2 ds \right]^{1/2} \\ &= O\left(\frac{1}{\sqrt{b_n}}\right) O\left(\frac{1}{d_n}\right),\end{aligned}\tag{6.17}$$

since  $\alpha_j$  is Lipschitz. The Lipschitz assumption on  $\alpha_j$  can also be used to show that

$$\sup_{t \in [0,1]} |\bar{\alpha}_j^{(n)}(t) - \alpha_j(t)| = O(b_n).\tag{6.18}$$

Combining (6.16) - (6.18) we obtain

$$\begin{aligned} \sup_{t \in [0,1]} |\tilde{\alpha}_j^{(n)}(t) - \alpha_j(t)| &= O_p(n^{\beta - \frac{1}{2}}) + O(n^{\frac{1}{2}\beta - \delta}) + O(n^{-\beta}) \\ &= O_p(n^{-\frac{1}{2}(1-2\beta)}), \end{aligned}$$

since  $\frac{1}{2}\beta - \delta \leq \frac{1}{2}\beta - \frac{1}{2}(1-\beta) = -\frac{1}{2}(1-2\beta)$  and  $-\beta \leq \beta - \frac{1}{2} = -\frac{1}{2}(1-2\beta)$  when  $\beta \geq \frac{1}{4}$ .

This completes the proof of the theorem.  $\square$

Acknowledgement. I would like to thank Fred Huffer for his helpful comments and interest in this work.

## REFERENCES

- Aalen, O.O. (1975). Statistical inference for a family of counting processes. Ph.D. Dissertation, Department of Statistics, University of California, Berkeley.
- Aalen, O.O. (1978). Nonparametric inference for a family of counting processes. Ann. Statist. 6, 701-726.
- Aalen, O.O. (1980). A model for nonparametric regression analysis of counting processes. Lecture Notes in Statistics 2, 1-25. Springer-Verlag, New York.
- Andersen, P.K. and Gill, R.D. (1982). Cox's regression model for counting processes: a large sample study. Ann. Statist. 10, 1100-1120.
- Arjas, E. (1986). A graphical method for assessing goodness-of-fit in Cox's proportional hazards model. Preprint.
- Billingsley, P. (1968). Convergence of Probability Measures. Wiley, New York.
- Breslow, N.E. (1986). Cohort analysis in epidemiology. In A.C. Atkinson and S.E. Fienberg (eds.), A Celebration of Statistics: the ISI Centenary Volume. Springer-Verlag, New York, 109-143.
- Buckley, J.D. (1984). Additive and multiplicative models for relative survival rates. Biometrics 40, 51-62.
- Chow, W.S. and Teicher, H. (1978). Probability Theory: Independence, Interchangeability, Martingales. Springer-Verlag, New York.
- Christopeit, N. (1986). Quasi-least-squares estimation in semimartingale regression models. Stochastics 16, 255-278.
- Cox, D.R. (1972). Regression models with life tables (with discussion). J. Roy. Statist. Soc. B34, 187-220.
- Dellacherie, C. (1972). Capacités et Processus Stochastiques. Springer-Verlag, Berlin.
- Dellacherie, C. and Meyer, P.-A. (1982). Probabilities and Potential B. North-Holland, Amsterdam.
- Grenander, U. (1981). Abstract Inference. Wiley, New York.
- Hall, W.J. and Wellner, J.A. (1980). Confidence bands for a survival curve with censored data. Biometrika 67, 133-143.
- Karr, A.F. (1983). Maximum likelihood estimation in the multiplicative intensity model. Center for Stochastic Processes, Department of Statistics, U.N.C. Chapel Hill, Technical report #46.

- Kopp, P.E. (1984). Martingales and Stochastic Integrals. Cambridge University Press.
- Le Breton, A. and Musiela, M. (1986). Order of convergence of least squares estimates in multiple linear regression. C.R. Acad. Sc. Paris Ser. I, 302, 333-336.
- Liptser, R.S. and Shiriyayev, A.N. (1978). Statistics of Random Processes, Volume II. Springer-Verlag, New York.
- McKeague, I.W. (1986a). Estimation for a semimartingale regression model using the method of sieves. Ann. Statist. 14, 579-589.
- McKeague, I.W. (1986b). Asymptotic theory for seive estimators in semimartingale regression models. Submitted to Ann. Statist.
- Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions. Ann. Statist. 11, 453-466.
- Rao, R.R. (1963). The law of large numbers for  $D[0,1]$ -valued random variables. Theor. Probab. Appl. 8, 70-74.



REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER FSU M741 USARO D-95	2. GOVT ACCESSION NO. <b>ADA173498</b>	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle)  Nonparametric Inference in Additive Risk Models for Counting Processes		5. TYPE OF REPORT & PERIOD COVERED  Technical Report
		6. PERFORMING ORG. REPORT NUMBER FSU Statistics Report M741
7. AUTHOR(s) Ian W. McKeague		8. CONTRACT OR GRANT NUMBER(s) DAAL03-86-K-0094
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Florida State University Tallahassee, FL 32306-3033		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Research Office - Durham P.O. Box 12211 Research Triangle Park, NC 27709		12. REPORT DATE August, 1986
		13. NUMBER OF PAGES 22
		15. SECURITY CLASS. (of this report)  Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report)  for public release; distribution unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)  Counting process, survival analysis, martingale methods, functional central limit theorem, regression models		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)  Nonparametric estimators for the hazard functions in an additive risk model for counting processes are studied. We establish a functional central limit theorem for the integrated estimators and show how this can be used to find the asymptotic null distribution of a maximal deviation statistic for Kolmogorov-Smirnov type testing. In addition, we provide confidence bands for approximations to the integrated hazard functions and show that certain smoothed versions of the hazard function estimators are uniformly consistent.		

END

12-86

DTIC